

**PEMANFAATAN MACHINE LEARNING DALAM ANALISIS
SENTIMEN DI APLIKASI TWITTER (X) : STUDI KASUS ISU
KABUR AJA DULU MENGGUNAKAN METODE SVM**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



Diajukan oleh

LINA ROHMATUN

21.11.4299

Kepada

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PEMANFAATAN MACHINE LEARNING DALAM ANALISIS
SENTIMEN DI APLIKASI TWITTER (X) : STUDI KASUS ISU
KABUR AJA DULU MENGGUNAKAN METODE SVM**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



Diajukan oleh

LINA ROHMATUN

21.11.4299

Kepada

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

**SKRIPSI
PEMANFAATAN MACHINE LEARNING DALAM ANALISIS SENTIMEN
DI APLIKASI TWITTER (X) : STUDI KASUS ISU KABUR AJA DULU
MENGUNAKAN METODE SVM**

yang disusun dan diajukan oleh
Lina Rohmatun
21.11.4299

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Agustus 2025

Dosen Pembimbing,



Anna Erlita, M.Kom
NIK. 190302290

HALAMAN PENGESAHAN
SKRIPSI
PEMANFAATAN MACHINE LEARNING DALAM ANALISIS SENTIMEN
DI APLIKASI TWITTER (X) : STUDI KASUS ISU KABUR AJA DULU
MENGGUNAKAN METODE SVM

yang disusun dan diajukan oleh

Lina Rohmatun

21.11.4299

Telah dipertahankan di depan Dewan Penguji

pada tanggal 19 Agustus 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Theonilus Bayu Sasongko, S.Kom., M.Eng
NIK. 190302375

Majid Rahardi, S.Kom., M.Eng
NIK. 190302393

Anna Baita, M.Kom
NIK. 190302290



Skripsi ini telah diterima sebagai salah satu persyaratan

untuk memperoleh gelar Sarjana Komputer

Tanggal 19 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama Mahasiswa : Lina Rohmatun
NIM : 21.11.4299

Menyatakan bahwa Skripsi dengan judul berikut:

**Pemanfaatan Machine Learning dalam Analisis Sentimen di Aplikasi Twitter
(X) : Studi Kasus Isu Kabur Aja Dulu menggunakan Metode SVM**

Dosen Pembimbing: Anna Baita, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Agustus 2025

Yang Menyatakan,



Lina Rohmatun

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam kehadiran Allah SWT, yang telah memberikan kekuatan, kesehatan, dan kelapangan waktu, skripsi ini penulis persembahkan kepada:

1. Bapak Fauzi dan Ibu Nur Hidayah tercinta, terima kasih atas doa yang tak pernah putus, kasih sayang yang tulus, perhatian yang penuh cinta, serta pengorbanan yang tidak pernah berhenti. Dengan penuh rasa syukur, saya persembahkan karya ini sebagai wujud cinta dan terima kasih yang mendalam. Semoga Ibu dan Bapak senantiasa diberi kesehatan, umur panjang, keberkahan, dan kebahagiaan dalam setiap langkah hidup. Adik tersayang, yang telah memberikan doa dan dukungan. Terima kasih telah menjadi penyemangat yang tak tergantikan.
2. Keluarga besar tersayang, yang berperan besar dalam memberikan semangat dan doa restu dalam setiap langkah perjalanan ini.
3. Sahabat penulis sejak semester satu, Fitria, Farhan, Anugrah, dan Daffa, terima kasih atas persahabatan yang penuh kasih sayang, canda tawa yang menguatkan, serta kebersamaan yang menjadi warna indah dalam setiap langkah perjalanan ini.
4. Teman-teman penulis dari waktu SMP hingga sekarang Nida, Arissa, Yulia yang sudah menemani dan berbagi dalam suka maupun duka. Terima kasih atas persahabatan yang penuh kasih sayang, canda tawa yang menguatkan, serta kebersamaan yang menjadi warna indah dalam setiap langkah perjalanan ini.
5. Almater tercinta, Universitas Amikom Yogyakarta, tempat penulis menimba ilmu, membentuk karakter, dan mengembangkan potensi diri.

Skripsi ini adalah bukti bahwa ketekunan, doa, dan semangat pantang menyerah akan selalu menemukan jalannya. Semoga skripsi ini dapat memberi manfaat, baik bagi penulis sendiri maupun bagi pembaca di masa yang akan datang.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“Pemanfaatan Machine Learning dalam Analisis Sentimen di Aplikasi Twitter (X): Studi Kasus Isu Kabur Aja Dulu Menggunakan Metode SVM”**. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Keterbatasan dalam hal kemampuan, pengetahuan, serta pengalaman menyebabkan masih adanya kekurangan yang perlu diperbaiki. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun agar karya tulis di masa mendatang dapat lebih baik lagi.

Proses penyusunan skripsi ini tidak terlepas dari berbagai kendala. Namun, berkat pertolongan Allah SWT, serta dukungan, bantuan, dan kerja sama dari berbagai pihak, baik secara moril maupun materil, akhirnya skripsi ini dapat terselesaikan dengan baik. Untuk itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom, selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
4. Ibu Anna Baita, M.Kom, selaku dosen pembimbing yang dengan penuh kesabaran dan ketulusan telah memberikan bimbingan, arahan, serta motivasi yang sangat berharga dalam penyusunan skripsi ini.

5. Seluruh dosen Fakultas Ilmu Komputer Program Studi Informatika Universitas Amikom Yogyakarta yang telah berbagi ilmu, wawasan, serta pengalaman berharga kepada penulis selama perkuliahan.
6. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan doa, semangat, dan dukungan sehingga penulis dapat menyelesaikan skripsi ini.

Akhir kata, penulis berdoa semoga Allah SWT memberikan balasan kebaikan yang berlipat ganda kepada semua pihak yang telah membantu. Penulis juga berharap skripsi ini dapat memberikan manfaat, baik bagi penulis sendiri maupun bagi para pembaca yang tertarik dengan bidang penelitian ini.

Yogyakarta, 16 Agustus 2025

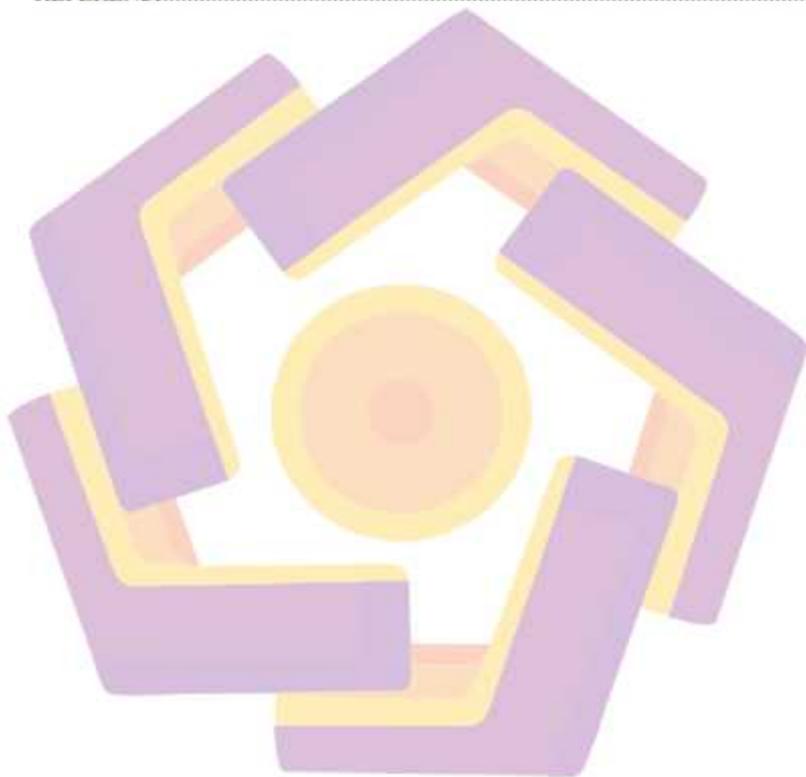
Penulis

DAFTAR ISI

HALAMAN JUDUL	1
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xi
BAB I PENDAHULUAN	xiii
1.1 Latar Belakang.....	5
1.2 Rumusan Masalah	6
1.3 Batasan Masalah.....	6
1.4 Tujuan Penelitian.....	7
1.5 Manfaat Penelitian.....	7
1.6 Sistematika Penulisan.....	8
BAB II TINJAUAN PUSTAKA	9
2.1 Studi Literatur.....	9
2.2 Dasar Teori.....	15
2.2.1 Twitter	15
2.2.2 Analisis Sentimen	15
2.2.3 <i>Data Scraping</i>	15
2.2.4 <i>Preprocessing Data</i>	16
2.2.5 <i>Labeling</i>	17
2.2.6 <i>Exploratory Data Analysis (EDA)</i>	18
2.2.7 <i>TF-IDF (Term Frequency - Inverse Document Frequency)</i>	19
2.2.8 <i>SMOTE (Synthetic Minority Oversampling Technique)</i>	19
2.2.9 <i>Support Vector Machine (SVM)</i>	20
2.2.10 <i>K-Fold Cross Validation</i>	22
2.2.11 <i>Hyperparameter Tuning</i>	23
2.2.12 Evaluasi Model.....	24
BAB III METODE PENELITIAN	26
3.1 Objek Penelitian	26

3.2	Alur Penelitian.....	26
3.2.1	<i>Scraping Data</i>	27
3.2.2	<i>Pre-processing Data</i>	28
3.2.3	<i>Labeling</i>	30
3.2.4	<i>Exploratory Data Analysis (EDA)</i>	31
3.2.5	<i>Feature Extraction (TF-IDF)</i>	31
3.2.6	<i>Split Data</i>	32
3.2.7	<i>SMOTE (Synthetic Minority Over-sampling Technique)</i>	32
3.2.8	<i>Modeling SVM (Support Vector Machine)</i>	33
3.2.9	<i>Validasi Model menggunakan K-Fold Cross-Validation</i>	33
3.2.10	<i>Hyperparameter Tuning (GridSearch)</i>	34
3.2.11	<i>Evaluasi Model</i>	34
3.3	Alat dan Bahan.....	35
3.3.1	Data Penelitian.....	35
3.3.2	Alat/Instrumen.....	35
BAB IV HASIL DAN PEMBAHASAN		37
4.1	Pengumpulan Data.....	37
4.2	Preprocessing Data.....	38
4.2.1	Mengecek jumlah nilai kosong (missing value).....	38
4.2.2	Menghapus Kolom Yang Tidak Dipakai.....	39
4.2.3	Mengecek Data Duplikat.....	40
4.2.4	Menghapus Link, Emoji, Simbol, Mention.....	40
4.2.5	<i>Case Folding</i>	41
4.2.6	<i>Tokenization</i>	41
4.2.7	<i>Stopword Removal</i>	42
4.2.8	<i>Remove Slang</i>	43
4.2.9	<i>Stemming</i>	44
4.2.10	<i>Translation</i>	44
4.3	<i>Labelling</i>	45
4.4	<i>Exploring Data Analysis (EDA)</i>	47
4.5	<i>Feature Extraction (TF-IDF)</i>	50
4.6	<i>Split Data</i>	51
4.7	<i>SMOTE (Synthetic Minority Over-sampling Technique)</i>	52

4.8	<i>Modelling Support Vector Machine (SVM)</i>	52
4.9	K-Fold Cross-Validation	53
4.10	<i>Hyperparameter Tunning (GridSearch)</i>	54
4.11	Evaluasi Model.....	55
BAB V PENUTUP		58
5.1	Kesimpulan.....	58
REFERENSI.....		60



DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	12
Tabel 4. 1. Informasi Dataset 37	
Tabel 4. 2 Hasil Cek Missing Value	39
Tabel 4. 3 Hasil Kolom Full_text	40
Tabel 4. 4 Hasil Proses Clean Text.....	40
Tabel 4. 5 Hasil Proses Case Folding	41
Tabel 4. 6 Hasil Proses Tokenization.....	42
Tabel 4. 7 Hasil Proses Stopword Removal	42
Tabel 4. 8 Hasil Proses Remove Slang	43
Tabel 4. 9 Hasil Proses Stemming	44
Tabel 4. 10 Hasil Proses Translation	45
Tabel 4. 11 Hasil Proses Pelabelan	46
Tabel 4. 12 Hasil Proses TF-IDF	51
Tabel 4. 13 Hasil Proses Modelling SVM.....	53
Tabel 4. 14 Kombinasi Parameter.....	54
Tabel 4. 15 Hasil Best Parameter Svm Linear.....	54
Tabel 4. 16 Hasil Best Parameter SVM Polynomial.....	55
Tabel 4. 17 Hasil Best Parameter SVM Sigmoid	55

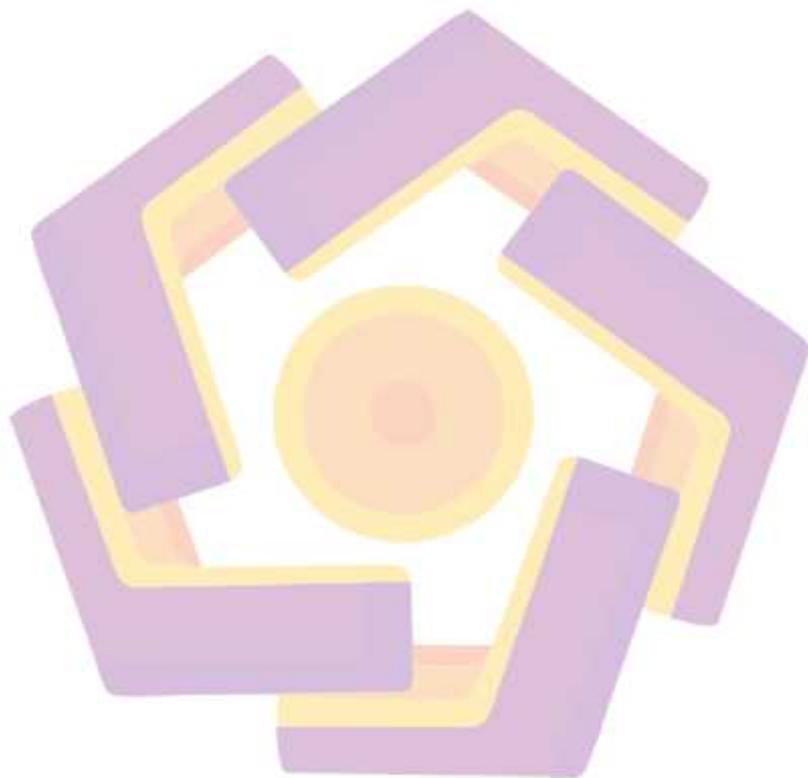
DAFTAR GAMBAR

Gambar 2. 2 Algoritma SVM	21
Gambar 2. 1 Ilustrasi K-Fold Cross Validation	23
Gambar 2. 3 Confusion Matrix.....	24
Gambar 3. 1 Alur Penelitian	27
Gambar 4. 1 Visualisasi Hasil Proses Labelling	
Gambar 4. 2 Hasil WordCloud Positif.....	48
Gambar 4. 3 Hasil WordCloud Negatif.....	49
Gambar 4. 4 Hasil WordCloud Netral	50
Gambar 4. 5 Hasil Proses Sebelum dan Sesudah SMOTE.....	52
Gambar 4. 6 Visualisasi Hasil Setiap Kernel.....	56
Gambar 4. 7 Hasil Confusion Matrix Kernel Terbaik.....	56

DAFTAR LAMPIRAN

Lampiran 1 source code dan data

Lampiran 2 LoA publikasi Bersama JAIC



DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machine
SMOTE	Synthetic Minority Oversampling Technique.
TF-IDF	Term Frequency - Inverse Document Frequency
NLP	Natural Language Processing
FN	False Negatif
FP	False Positif
TN	True Negatif
TP	True Positif
CSV	Comma Separated Values
NLTK	Natural Language Toolkit

DAFTAR ISTILAH

Sentimen	Ekspresi perasaan, opini, atau evaluasi pengguna terhadap suatu isu, produk, atau layanan, yang biasanya dikategorikan sebagai positif, negatif, atau netral.
Imbalance Data	Situasi di mana jumlah sampel antar kelas tidak proporsional, sehingga dapat memengaruhi kinerja model prediksi.
Optimasi	Upaya sistematis untuk menemukan kombinasi parameter atau strategi terbaik agar model bekerja secara maksimal.
Oversampling	Teknik memperbanyak sampel pada kelas yang jarang muncul, sehingga distribusi data antar kelas menjadi lebih seimbang.
Hyperparameter	Parameter yang ditetapkan sebelum proses pelatihan model machine learning dimulai.
Akurasi	Rasio antara jumlah prediksi yang benar dengan keseluruhan jumlah data yang diuji.
Precision	Proporsi data yang benar-benar relevan di antara semua data yang diprediksi relevan oleh model.
Recall	Proporsi data relevan yang berhasil ditemukan oleh model dibandingkan dengan seluruh data relevan yang ada.
F1-score	Ukuran kinerja model yang menggabungkan precision dan recall secara seimbang melalui rata-rata harmonis.
Kernel	Fungsi transformasi yang digunakan dalam SVM untuk memetakan data dari ruang berdimensi rendah ke ruang berdimensi lebih tinggi agar lebih mudah dipisahkan.
C	Parameter yang mengatur tingkat toleransi terhadap kesalahan klasifikasi, sehingga memengaruhi kompleksitas dan margin model.
Gamma	Parameter yang menentukan seberapa jauh pengaruh satu titik data meluas ke titik lain dalam ruang fitur.
Degree	Besaran pangkat yang digunakan dalam kernel polynomial untuk menentukan kompleksitas fungsi pemisah.

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap isu #KaburAjaDulu di Twitter (X) menggunakan metode Support Vector Machine (SVM). Data diperoleh melalui tahapan preprocessing dan representasi fitur dengan Term Frequency–Inverse Document Frequency (TF-IDF). Untuk menangani ketidakseimbangan data, digunakan teknik Synthetic Minority Oversampling Technique (SMOTE). Evaluasi model dilakukan dengan k-fold cross validation untuk memperoleh fold terbaik, yang kemudian digunakan pada tahap lanjutan perbandingan kernel dan optimasi. Model SVM dengan kernel Linear, Polynomial, dan Sigmoid dibandingkan pada berbagai rasio data latih-uji, serta dilakukan optimasi hyperparameter tuning menggunakan GridSearch. Hasil pengujian menunjukkan bahwa pada rasio 80:20, kernel Sigmoid memberikan performa terbaik dengan akurasi 0.7804, precision 0.7194, dan f1-score 0.7154. Kernel Linear juga cukup kompetitif dengan akurasi 0.7597, sedangkan kernel Polynomial menghasilkan akurasi lebih rendah yaitu 0,6433. Hasil GridSearch menunjukkan parameter optimal berbeda untuk setiap kernel, yaitu Linear ($C=1$), Polynomial ($C=1$, Degree=2, Gamma=scale), dan Sigmoid ($C=1$, Gamma=scale). Proses tuning terbukti meningkatkan performa Polynomial (akurasi naik dari 0,6902 menjadi 0.7108), namun tidak memberikan hasil lebih baik pada Linear maupun Sigmoid. Dengan demikian, kernel Sigmoid pada rasio 80:20 dapat dianggap sebagai model paling optimal dalam penelitian ini.

Kata kunci : Analisis Sentimen, *Twitter (X)*, *Support Vector Machine (SVM)*, Kabur Aja Dulu

ABSTRACT

This study aims to analyze public sentiment toward the #KaburAjaDulu issue on Twitter (X) using the Support Vector Machine (SVM) method. Data was obtained through preprocessing and feature representation using Term Frequency–Inverse Document Frequency (TF-IDF). To address data imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was employed. Model evaluation was conducted using k -fold cross-validation to identify the optimal fold, which was then utilized in subsequent kernel comparison and optimization stages. SVM models with Linear, Polynomial, and Sigmoid kernels were compared across various training-to-test data ratios, and hyperparameter tuning optimization was performed using GridSearch. The test results show that at an 80:20 ratio, the Sigmoid kernel provides the best performance with an accuracy of 0.7815, precision of 0.778, and $f1$ -score of 0.7789. The Linear kernel is also quite competitive with an accuracy of 0.7804, while the Polynomial kernel produces a lower accuracy of 0.6433. The GridSearch results show that the optimal parameters differ for each kernel, namely Linear ($C=1$), Polynomial ($C=1$, Degree=2, Gamma=scale), and Sigmoid ($C=1$, Gamma=scale). The tuning process improved the performance of the Polynomial kernel (accuracy increased from 0.6902 to 0.7804), but did not yield better results for the Linear or Sigmoid kernels. Therefore, the Sigmoid kernel at an 80:20 ratio can be considered the most optimal model in this study.

Keyword : Sentiment Analysis, Twitter, Support Vector Machine (SVM), Kabur Aja Dulu